Naive Bayes * we want to return probabilities and classify using the class with the highest probability

naive bayes classifier

Bayes Rule: $P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$

$P(B) = P(B|A)P(A) + P(B|\text{not } A)P(\text{not } A)$

Naive Bayes Assumption
- to make calculations easier
- assume that X_i and X_j are conditionally independent given Y for all $i \neq j$

$$P(X_1, X_2, \dots, X_d | Y) = \prod_{i=1}^{d} P(X_i | Y)$$

Conditional Independence

$$P(X|Y,Z) = P(X|Z)$$

Naive Bayes Classifier

$$\hat{y} = \text{argmax}_y \, P(Y=y | X=x)$$

apply Bayes rule and ignore denominator

$$= \text{argmax}_y \, P(Y=y) P(X=x | Y=y)$$

naive bayes assumption

$$= \text{argmax}_y \, P(Y=y) \prod_{i=1}^{d} P(X_i = x_i | Y=y)$$

in naive bayes, we don't have to do optimization for the parameter.

rather, we try different values of y and use the y with the highest probability

---

Maximum Likelihood Estimation
- given a sample of data, find the parameter that maximize probability of seeing the data
- normally take log, since it is easier to get derivative of logs and exponents become products
- take derivative and set to 0

I observed 5 Heads and 23 Tail. Find θ parameter of Heads probability.

$$P_\theta(D) = \theta^5 (1-\theta)^{23}$$
$$\log(P_\theta(D)) = \log(\theta^5 (1-\theta)^{23})$$
$$= \log(\theta)^5 + \log(1-\theta)^{23}$$
$$= 5\log(\theta) + 23\log(1-\theta)$$
$$\log(P_\theta(D)) = 5\log(\theta) + 23\log(1-\theta)$$
$$\frac{\partial}{\partial\theta}[\log(P_\theta(D))] = \frac{\partial}{\partial\theta}[5\log(\theta) + 23(1-\theta)] = 0$$
$$0 = \frac{5}{\theta} - \frac{23}{1-\theta}$$
$$\frac{23}{1-\theta} = \frac{5}{\theta}$$
$$23\theta = 5 - 5\theta$$
$$28\theta = 5$$
$$\theta = \frac{5}{28}$$